Army Research Laboratory

Validation Procedures for Equations, Algorithms, and Submodels

by Linda L. C. Moss

ARL-TN-245 September 2005

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5068

ARL-TN-245 **September 2005**

Validation Procedures for Equations, Algorithms, and Submodels

Linda L. C. Moss
Survivability/Lethality Analysis Directorate, ARL

REPORT DOCUMENTATION PAGE					*Form Approved* *OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**					
1. REPORT DATE *(DD-MM-YYYY)* September 2005	**2. REPORT TYPE** Final				**3. DATES COVERED (From - To)** February–May 2004
4. TITLE AND SUBTITLE Validation Procedures for Equations, Algorithms, and Submodels					**5a. CONTRACT NUMBER**
					5b. GRANT NUMBER
					5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Linda L. C. Moss					**5d. PROJECT NUMBER** 622618.H80
					5e. TASK NUMBER
					5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory ATTN: AMSRD-ARL-SL-BD Aberdeen Proving Ground, MD 21005-5068					**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TN-245
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					**10. SPONSOR/MONITOR'S ACRONYM(S)**
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Verification and validation is a continuous activity throughout the entire modeling and simulation life cycle. This document provides a template to address general validation procedures for equations, algorithms, and submodels that form the underlying analysis found in vulnerability/lethality software tools. This template is considered a bottom-up approach for validation of a submodel as a stand-alone model as opposed to a top-down approach for validation of the overall model for live-fire tests.					
15. SUBJECT TERMS verification, validation, sensitivity analysis, goodness-of-fit					
16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Linda L. C. Moss
a. REPORT UNCLASSIFIED	**b. ABSTRACT** UNCLASSIFIED	**c. THIS PAGE** UNCLASSIFIED	UL	20	**19b. TELEPHONE NUMBER** (*Include area code*) 410-278-6513

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

INTENTIONALLY LEFT BLANK.

1. Purpose

Verification and validation (V&V) is a continuous activity throughout the entire modeling and simulation (M&S) life cycle (Balci, 1998; Harmon and Youngblood, 1998), and "there is no single step-by-step checklist of tasks or events or a single method of V&V that will apply for every M&S as it goes through its [life cycle management]" (Department of Defense [DOD] Pamphlet 5–11, 1999). However, according to the DOD directive 5000.59 (1994), DOD components, such as the U.S. Army Research Laboratory, shall establish verification, validation, and accreditation (VV&A) policies and procedures for M&S applications.

This document provides a template to address general validation procedures for equations, algorithms, and submodels that form the underlying analysis found in vulnerability/lethality (V/L) software tools such as the Survivability/Lethality Analysis Directorate's M&S tool called MUVES. This template is considered a bottom-up approach for validation of a submodel as a stand-alone model, as opposed to a top-down approach for validation of the overall model for live-fire tests (Deitz et al., 1996). Logical and code verification that include computer-aided software engineering tools are not addressed here, but need to be included as an overall V&V process for software accreditation.

2. Definitions

The definitions of VV&A are given as defined in DOD 5000.59 and DOD Pamphlet 5–11. The interrelationship of verification and validation is illustrated in figure 1 (Deitz et al., 1996).

Verification: The process of determining that a model implementation accurately represents the developer's conceptual description and specification. Verification also evaluates the extent to which the model or simulation has been developed using sound and established software-engineering techniques.

Validation: The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. Validation methods include expert consensus, comparison with historical results, comparison with test data, peer review, and independent review.

Accreditation: The official certification that a model or simulation is acceptable for use for a specific purpose.

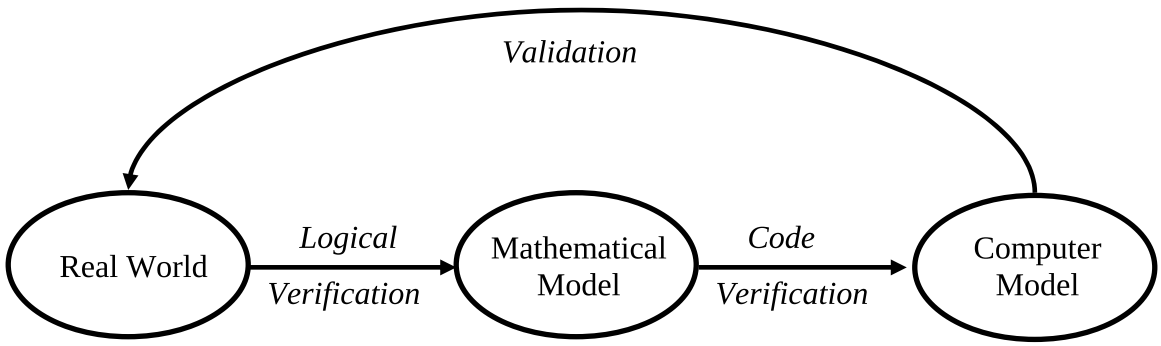

Figure 1. Role of verification and validation (Deitz, 1996).

3. Summary of Model Validation

Saucier and Baker (1996) summarized model validation in the following way (see also Deitz et al., 1996; DOD Pamphlet 5–11, 1999; Knepell and Arango, 1993).

3.1 Face Validation

Face validation is performed by subject-matter experts (SMEs) who check that the model produces reasonable results for the intended application based upon its performance. This is a point of departure to determine courses of action for more comprehensive validation efforts. This is a subjective evaluation, yet it can identify areas of potential weaknesses or strengths in the model, is low cost, and has a short timeframe for completion.

3.2 Model Applicability, Assumptions, and Limitations

Model applicability, assumptions, and limitations are checked to make sure that they are appropriate to the phenomena they represent.

3.3 Data Audit and Availability

Data audit and availability assures that the data collection techniques are consistent and well-documented and that the data requirements for the model are realistic.

3.4 Comparison to Laboratory and Operational Data

Submodels are compared to results obtained under laboratory conditions. The overall model is also compared to laboratory conditions, where possible, as well as field experiments (operational data).

In the effort to address the more comprehensive validation process that needs to be considered throughout M&S development, a template is provided in section 4. The template specifies statistical quantitative measures of model fits and validation data fits, as well as graphical comparisons. It is not intended to replace reports that would provide details of the experimental program and model development, but to be used to guide such reports and identify areas of future research or methodology by identifying limitations, trade-offs, unresolved issues, and suggested improvements.

4. Validation Template for Equations, Algorithms, and Submodels

Model Name:

Provide model name.

Version (or Date):

Provide version or date of the model. Indicate if it is an update to a previous version.

Purpose and Uses:

Define the purpose of the model and identify its major application.

Model Description:

Provide the mathematical model (the equation or set of equations) describing the phenomenology.

Model Inputs:

List and describe the required model inputs.

Model Outputs:

List and describe the model output(s).

Range of Applicability:

Provide the range of applicability or boundary conditions for each input variable and output metric (or distribution of the inputs and outputs).

Assumptions:

List the underlying theoretical or empirical foundation and assumptions. Specify how each is appropriate to the phenomena that it represents.

Evaluation Criteria:

Describe the real-world data (live-fire data, component data, test data, operational data, etc.) with which the model will be compared. Provide experimental data or references of the comparable data. If known, provide confidence levels, reliability/probability levels, ballistic protection levels, etc. from which the goodness-of-fit will be evaluated.

Considerations in the Model Building Process:

Verification and validation are on-going activities for establishing credibility throughout the entire modeling process from algorithm development to testing with confirmation data. These activities provide documentation for the accreditation process. Careful considerations of *data collection and quantification of the equations to the data in the early stages of model development can save resources in the latter stages. Therefore, it is recommended that statistical experimental design techniques be considered when building empirical or semi-empirical models. Tools to quantify model fits are given in the Goodness-of-Fit section. These tools in graphical methods and statistical methods apply to data used for development, as well as to data used to validate an existing model.*

A. Screening Designs to Identify Key Model Parameters—*When building an empirical or semi-empirical model, it may be necessary because of cost or time constraints to conduct small experiments using statistical design methods to determine which of the potential explanatory variables are the key parameters in the model. Such designs could include a 2^n factorial or fractional factorials, Latin square, randomized block, split plot, etc.* (Hicks and Turner, 1999; Kirk, 1982; Montgomery, 1991). *To determine if variables are statistically significant, use regression analysis, analysis of variance, T-tests, or other statistical methods* (Draper and Smith, 1981; Kirk, 1982; Myers, 1986). *Screening designs are used to narrow the scope of the potential explanatory variables considered for the model. They typically do not address interaction among the variables.*

B. Experimental Design for Model Development—*Conduct a larger experimental program using the significant factors identified in the screening design to collect data from which empirical or semi-empirical models will be constructed. As provided in the description of the screening designs, statistical design techniques can be used to aid in this process. The range of each factor should extend to the expected real-world domain, at a minimum select a low, a medium, and a high value of each variable of interest. Including mid-range values along with the low and high values allows for building curvature in the model. Theoretical models may not have data for building the model; however, data will be required for validation.*

Confirmation Experiments (Validation Data):

After the model has been built, conduct additional experiments to collect new data that will be used to validate the model. List data used in the validation effort or provide references (reports) of the data. This data would not have been used in the development of the algorithm. For example, it may be a different projectile type or propellant type that was not used in model formulation. The data could have been collected at the same time as the data for model development but set aside for validation. Provide range of each input variable. This range should cover the bounds of the real-world data or subset of data for areas of the unvalidated domain. Use statistical experimental design techniques where possible.

Goodness-of-Fit:

A. Graphical Methods: *An easy yet important method of evaluating the goodness-of-fit of the model is graphing the model fit* (figure 2a). *Plot curves of each output variable vs. input variable(s) as in, say, Y vs. X or V_r vs. V_s curves. Show plot of predicted output vs. observed* *output* (figure 2b). *A diagonal line will show the perfect-fit line. Points above the line overpredict the response variable; points below the line underpredict. Plotting standardized residuals against each of the inputs helps identify areas that do not fit the response well. The adjusted R-sq (R^2) and standard error (s.e.) given in the figure are defined in the statistical methods section that follows.*

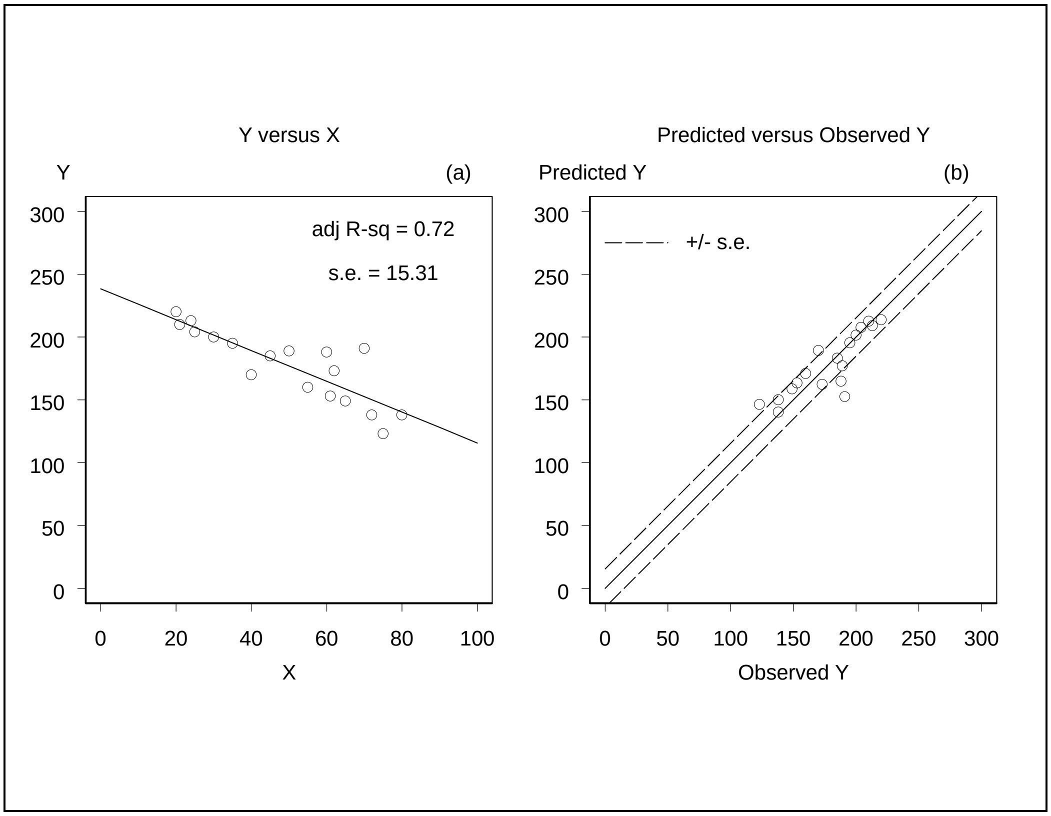

Figure 2. Y vs. X plot and predicted vs. observed plot for output variable Y.

Plot the confirmation/validation data along with the experimental data and model fit. Again, provide a predicted vs. observed plot of the output variable(s) (figure 2b) or output variable(s) against input variable(s) (e.g., V_r vs. V_s curve). Describe how the validation data fits within the distribution of the model. This can be in the form of box-plots or histograms. If confidence intervals were computed for the model, compare validation data with the confidence interval(s). Note that if the validation data do not fit well, additional modeling efforts may be required and another series of validation data may need to be generated before the model can be formally validated.

B. Statistical Methods: *Report standard error (the square root of the model variance. In a regression analysis, it is the square root of the residual mean square* $= \sqrt{\sum (predicted - observed)^2 / df}$ *, where df is the number of degrees of freedom for the residual error; i.e., the number of data points minus the number of coefficients estimated for the* *model.) Report the adjusted square of the correlation coefficient (R^2) or the adjusted R^2 fit of the data.*[1] *During model development, the statistically significant variables can be determined by analysis of variance or regression analysis. Report the p-values for each. Develop confidence intervals where possible for the model fit and each coefficient. Identify any nuisance variables: these are other sources of variability (measurement error, use of different operators, gunners, etc.) that may contribute to differences in response. Sometimes these nuisance effects can be separated from the effects of the main model parameters in a statistical experimental design, such as blocking in a randomized block design.*

Sensitivity Analysis:[2]

Vary each input variable (e.g., +/-25% [Collins, 2003] *) both individually and simultaneously to determine the change in the response variable(s). Test at boundary conditions (data at the extreme values) to stress the model to real-world application. If statistical experimental designs are used in the modeling phase, each variable should be statistically significant; however, the amount of change in the response variable may not have a practical significance depending on the V/L analysis conducted for the customer. These analyses may not be known at the time of model development, but sensitivity studies will quantify the impact of changes to the inputs and then further guide subsequent V/L analyses.*

Limitations and Trade-Offs:

List model limitations or known deficiencies. Provide information on the trade-offs of using one empirical model vs. another or using an empirical model vs. a physics-based model (e.g., computational fluid dynamic codes take too long).

Unresolved Issues:

Describe any validation tests that resulted in anomalies.

Suggested Improvements and/or Extensions:

Provide recommendations to address the model limitations and any extensions to the model that would make it more robust.

Points of Contact:

Include the name(s) and specific areas of responsibility (such as developer, reviewer, SMEs) for clarification, additional information, or discussion.

References and/or Attachments:

Provide a list of references and any supporting documents as well as attachments that describe the model or the model fits to the data. Fits include graphical techniques and

---

[1] R^2 is the amount of variability in the data that is explained by the model. It is a value between 0 and 1, where 1 is a perfect fit. The adjusted R^2 accounts for the number of variables in the model so that a model with few variables to explain the phenomenon will get more credit than a model that requires many variables.

[2] Sensitivity analysis is also considered as a step in the verification process in addition to the validation process in DOD Pamphlet 5-11 (1999).

quantitative measures (see Goodness-of-Fit section). Include experimental data or list of documents that contain data used to develop the model.

5. Conclusion

The template provided in section 4 provides a convenient outline for documenting parameters in the development of equations, algorithms, and submodels that are part of an underlying vulnerability/lethality M&S. It is not only important in the V&V process to quantify the model fits to experimental data, but to also identify the range of applicability, assumptions for model use, known limitations, and any unresolved issues.

Further information in the VV&A process for M&S can be found in the DOD 5000.61 (2003) and Defense Modeling Simulation Office (2000).

6. References

Balci, O. Verification, Validation, and Accreditation. Medeiros, D. J., Watson, E. F., Carson, J. S., Manivannan, M. S., Eds.; *Proceedings of the 1998 Winter Simulation Conference,* Washington, DC, 13–16 December 1998.

Collins, J. C. *Sensitivity Analysis of MUVES/S2 to Behind-Armor Debris*; ARL-TR-3067; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2003.

Defense Modeling Simulation Office (DSMO). Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide. OSD/DDR&E. http://vva.dmso.mil, 16 May 2000.

Deitz, P. H.; Saucier, R.; Baker, W. Developments in Modeling Ballistic Live-Fire Events. *16th International Symposium on Ballistics,* San Francisco, CA, 23–28 September 1996.

Department of Defense. *DOD Modeling and Simulation (M&S) Management*; Number 5000.59, 4 January 1994.

Department of Defense. *DOD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)*; Number 5000.61, 13 May 2003.

Department of Defense. *Verification, Validation, and Accreditation of Army Models and Simulations*; Pamphlet 5-11; Headquarters, Department of the Army: Washington, DC, 30 September 1999.

Draper, N.; Smith, H. *Applied Regression Analysis;* 2nd ed.; John Wiley & Sons: New York, 1981.

Harmon, S.; Youngblood, S. Modeling Simulation Validation Process Maturity. *Procedures for Accreditation of Software by the Joint Technical Coordinating Group for Munitions Effectiveness*. Joint Technical Coordinating Group for Munitions Effectiveness, 61 JTCG/ME-1-7, 27 February 1998.

Hicks, C. R.; Turner, K. V., Jr. *Fundamental Concepts in the Design of Experiments*; 5th ed.; Oxford University Press, Inc., 1999.

Kirk, R. E. *Experimental Design: Procedures for the Behavioral Sciences*; 2nd ed.; Brooks/Cole Publishing Company: Belmont, CA, 1982.

Knepell, P. L.; Arangno, D. C. *Simulation Validation, A Confidence Assessment Methodology*; IEEE Computer Society Press: Los Alamitos, CA, 1993.

Montgomery, D. C. *Design and Analysis of Experiments*; 3rd ed.; John Wiley & Sons: New York, 1991.

Myers, R. H. *Classical and Modern Regression With Applications*; 2nd ed.; Duxbury Press, PWS-KENT Publishing Company: Belmont, CA, 1986.

Saucier, R.; Baker, W. E. Resource Requirements for VV&A of MUVES SQuASH. U.S. Army Research Laboratory, Survivability/Lethality Analysis Directorate, Ballistic Vulnerability/Lethality Division; Briefing to Mr. Hollis and Dr. Fallin, Pentagon, Washington, DC, 17 April 1996.

Bibliography

Baker, W. E.; Saucier, R.; Muehl, T. M.; Grote, R. L. *Comparison of MUVES-SQuASH With Bradley Fighting Vehicle Live-Fire Test Results*; ARL-TR-1846; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, November 1998.

Chambers, J. M.; Cleveland, W. S.; Kleiner, B.; Tukey, P. A. *Graphical Methods for Data Analysis*; Wadsworth & Brooks/Cole Publishing Company Advanced Books & Software: Pacific Grove, CA, 1983.

Collins, J. C. *Sensitivity Analysis in Testing the Consistency of Vulnerability Models*; ARL-TR-2668; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, February 2002.

Conover, W. J. *Practical Nonparametric Statistics*; 3rd ed.; John Wiley & Sons, Inc.: New York, 1999.

Deitz, P. H.; Saucier, R.; Baker, W. *Modeling Ballistic Live-Fire Events Trilogy*; ARL-TR-1274; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 1996.

Hunt, J. E., Dibelka, R. E.; Winner, W. A.; Milam, M. D.; Gillich, C. J. *MUVES-S2 Accreditation Support Package: Volume I*; ARL-TR-3025; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, July 2003.

Snyder, E. A.; Wilson, L. W.; Thompson, N. P.; Winner, W. A. *Verification and Validation of MUVES-S2 in Support of the BAT Lethality Assessment*; ARL-TR-3245; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, July 2004.

Tonnessen, L.; Fries, A.; Starkey, L.; Stein, A. *Live Fire Testing in the Evaluation of the Vulnerability of Armored Vehicles and Other Exposed Land-Based Systems*; IDA Paper P-2205, Institute for Defense Analysis, July 1989.

Turner, B. W.; Barr, B.; Okamoto, P. Y. *Independent Assessment of SQuASH*; briefing package presented to SLAD; Institute for Defense Analysis (IDA), 18 February 1994.

Webb, D. W. *Tests for Consistency of Vulnerability Models*; BRL-TR-3030; U.S. Army Ballistic Research Laboratory: Aberdeen Proving Ground, MD, August 1999.

NO. OF COPIES	ORGANIZATION
1 (PDF ONLY)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	US ARMY RSRCH DEV & ENGRG CMD SYSTEMS OF SYSTEMS INTEGRATION AMSRD SS T 6000 6TH ST STE 100 FORT BELVOIR VA 22060-5608
1	INST FOR ADVNCD TCHNLGY THE UNIV OF TEXAS AT AUSTIN 3925 W BRAKER LN STE 400 AUSTIN TX 78759-5316
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC IMS 2800 POWDER MILL RD ADELPHI MD 20783-1197
3	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197
3	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CS IS T 2800 POWDER MILL RD ADELPHI MD 20783-1197

ABERDEEN PROVING GROUND

1	DIR USARL AMSRD ARL CI OK TP (BLDG 4600)

NO. OF
COPIES ORGANIZATION

1 OUSD(AT)/S&T AIR WARFARE
R MUTZELBURG
RM 3E139
3090 DEFENSE PENTAGON
WASHINGTON DC 20301-3090

1 OUSD(AT)/S&T LAND WARFARE
A VIILU
RM 3B1060
3090 DEFENSE PENTAGON
WASHINGTON DC 20310-3090

1 UNDER SECY OF THE ARMY
DUSA OR
W HOLLIS
RM 2D564
102 ARMY PENTAGON
WASHINGTON DC 20310-0102

1 ASST SECY ARMY
ACQSTN LOGISTICS & TECH
SAAL ZP RM 2E661
103 ARMY PENTAGON
WASHINGTON DC 20310-0103

1 ASST SECY ARMY
ACQSTN LOGISTICS & TECH
SAAL ZS RM 3E448
103 ARMY PENTAGON
WASHINGTON DC 20310-0103

1 DIRECTOR FORCE DEV
DAPR FDZ
RM 3A522
460 ARMY PENTAGON
WASHINGTON DC 20310-0460

1 US ARMY TRADOC ANL CTR
ATRC W
A KEINTZ
WSMR NM 88002-5502

1 USARL
AMSRD ARL SL EA
R FLORES
WSMR NM 88002-5513

1 USARL
AMSRD ARL SL EI
J NOWAK
FORT MONMOUTH NJ 07703-5601

NO. OF
COPIES ORGANIZATION

ABERDEEN PROVING GROUND

1 US ARMY DEV TEST COM
CSTE DTC TT T
APG MD 21005-5055

1 US ARMY EVALUATION CTR
CSTE AEC SVE
R BOWEN
4120 SUSQUEHANNA AVE
APG MD 21005-3013

1 US ARMY EVALUATION CTR
CSTE AEC SVE S
R POLIMADEI
4120 SUSQUEHANNA AVE
APG MD 21005-3013

1 US ARMY EVALUATION CTR
CSTE AEC SV L
R LAUGHMAN
4120 SUSQUEHANNA AVE
APG MD 21005-3013

12 DIR USARL
AMSRD ARL SL
J BEILFUSS
P TANENBAUM
AMSRD ARL SL B
J FRANZ
M PERRY
AMSRD ARL SL BB
D BELY
D FARENWALD
S JUARASCIO
AMSRD ARL SL BD
R GROTE
AMSRD ARL SL BE
L ROACH
AMSRD ARL SL E
M STARKS
AMSRD ARL SL EC
J FEENEY
E PANUSKA

NO. OF
COPIES ORGANIZATION

ABERDEEN PROVING GROUND

1 US ARMY EVALUATION CTR
CSTE AEC SV L
R LAUGHMAN
4120 SUSQUEHANNA AVE
APG MD 21005-3013

49 DIR USARL
AMSRD ARL SL B
R SANDMEYER
AMSRD ARL SL BB
W WINNER
L WILSON
T MALLORY
R DIBELKA
G BRADLEY
D LYNCH
L LOSIE
M KUNKEL
P SMITH
J ROBERTSON
J SHORT
A PRICE
D JORDAN
R GANGLER
M BURDESHAW
J PLOSKONKA
P MERGLER
AMSRD ARL SL BD
J COLLINS
L MOSS (15 CPS)
F MARSH
S POLYAK
J POLESNE
B SMITH
E SNYDER
AMSRD ARL SL BE
J ABELL
J AUTEN
L BUTLER
E DAVISSON
P FEDELE
K BATES
M MAHAFFEY
W MERMAGEN
D PETTY
R SAUCIER

INTENTIONALLY LEFT BLANK.